VEGETATION CLASSIFICATION AND SURVEY

International Association for Vegetation Science (IAVS)

Vegetation Classification and Survey 3: 29–43
doi: 10.3897/VCS.67537

RESEARCH PAPER   INTERNATIONAL VEGETATION CLASSIFICATION

Integration of vegetation classification with land cover mapping: lessons from regional mapping efforts in the Americas

Patrick J. Comer[1], Jon C. Hak[2], Daryn Dockter[3], Jim Smith[4]

1 *Science Applications & Analysis, NatureServe, Boulder, CO, USA*
2 *private consultant, Longmont, CO, USA*
3 *KBR, Contractor to the U.S. Geological Survey (USGS) Earth Resources Observation and Science (EROS) Center, Sioux Falls, SD, USA*
4 *The Nature Conservancy, Jacksonville, FL, USA*

Corresponding author: Patrick J. Comer (pat_comer@natureserve.org)

Academic editor: Gonzalo Navarro-Sanchez ♦ **Received** 18 April 2021 ♦ **Accepted** 14 December 2021 ♦ **Published** 15 February 2022

Abstract

Aims: Natural resource management and biodiversity conservation rely on inventories of vegetation that span multiple management or political jurisdictions. However, while remote sensing data and analytical tools have enabled production of maps at increasing spatial resolution and reliability, there are limited examples where national or continental-scaled maps are produced to represent vegetation at high thematic detail. We illustrate two examples that have bridged the gap between traditional land cover mapping and modern vegetation classification. **Study area:** Our two case studies include national (USA) and continental (North and South America) vegetation and land cover mapping. These studies span conditions from subpolar to tropical latitudes of the Americas. **Methods:** Both case studies used a supervised modeling approach with the International Vegetation Classification (IVC) to produce maps that provide for greater thematic detail. Georeferenced locations for these vegetation types are used by machine learning algorithms to train a predictive model and generate a distribution map. **Results:** The USA LANDFIRE (Landscape Fire and Resource Management Planning Tools Project) case study illustrates how a history of vegetation-based classification and availability of key inputs can come together to generate standard map products covering more than 9.8 million km^2 that are unsurpassed anywhere in the world in terms of spatial and thematic resolution. That being said, it also remains clear that mapping at the thematic resolution of the IVC Group and finer resolution require very large and spatially balanced inputs of georeferenced samples. Even with extensive prior data collection efforts, these remain a key limitation. The NatureServe effort for the Americas - encompassing 22% of the global land surface - demonstrates methods and outputs suitable for worldwide application at continental scales. **Conclusions:** Continued collection of input data used in the case studies could enable mapping at these spatial and thematic resolutions around the globe.

Abbreviations: CART = Classification and Regression Tree; **CONUS** = Conterminous United States; **DSWE** = Dynamic Surface Water Extent; **EPA** = United States Environmental Protection Agency; **FGDC** = Federal Geographic Data Committee; **IVC** = International Vegetation Classification; **LANDFIRE** = Landscape Fire and Resource Management Planning Tools Project; **LFRDB** = LANDFIRE Reference Database; **LiDAR** = Light Detection and Ranging; **NDVI** = Normalized Difference Vegetation Index; **NLCD** = National Land Cover Database; **USNVC** = United States National Vegetation Classification; **USA** = United States of America; **WWF** = World Wildlife Fund or Worldwide Fund for Nature.

Keywords

America, distribution modeling, Random Forest, vegetation classification, vegetation mapping

Introduction

Natural resource management and biodiversity conservation often rely on inventories of vegetation that span multiple management or political jurisdictions. However, while remote sensing data and analytical tools have enabled production of maps at increasing spatial resolution and accuracy, there are limited examples where maps are produced at large national or continental scale to represent vegetation at high thematic detail. For example, the U.S. National Land Cover Database (NLCD) (Wickham et al. 2014) depicts land cover at 30 m pixel resolution, and is periodically repeated, enabling detection of major trends in land use of relevance to a broad range of resource management decisions. However, these and similar land cover products around the world depict relatively few distinct map classes (Wickham et al. 2021).

As vegetation classification has advanced, the potential to map far more ecologically distinct map classes presents important opportunities to address pressing societal needs (Lavrinenko 2020). Regional map products of increasingly high thematic and spatial resolution have proven essential for successfully mapping species ranges (Aycrigg et al. 2010), assessing ecosystem representation within protected areas (Pliscoff and Fuentes-Castillo 2011), and in systematic, place-based conservation planning (Groves and Game 2016). They are needed for documenting relative risk of ecosystem collapse, as can be documented through the International Union for Conservation of Nature (IUCN) Red List of Ecosystems framework (Keith et al. 2013). Additionally, depiction of vegetation structure and composition at moderate to finer levels of thematic detail enable assessment of dynamic processes, such as wildfire regimes (Rollins 2009).

We will review two case studies of both national and continental land cover mapping that have bridged the gap between traditional land cover mapping and modern vegetation classification. We trace recent development of terrestrial ecological classification in the Americas with specific reference to land cover mapping applications at regional to national and continental scales. This experience assisted in development of the International Vegetation Classification (IVC) (Faber-Langendoen et al. 2018). In turn, that classification hierarchy has been utilized directly in the mapping process.

The two case studies span conditions from subpolar to tropical latitudes of the Americas. One, limited to the United States, aimed to map the current distribution of the Group level of the IVC hierarchy (Table 1). The second aimed to map both potential/historical distributions of IVC Macrogroup Level of the IVC hierarchy (Table 1) across temperate and tropical North America, and all of South America.

Our first case study reviews the experience of the U.S. Landscape Fire and Resource Management Planning Tools Project (LANDFIRE) that, since the mid 2000s, has produced a series of moderate-to-high resolution national map products to facilitate strategic decision support to both wildfire and wildlife habitat managers. Resultant map layers describe vegetation composition and structure and can be compared to expected conditions to indicate alteration to expected natural wildfire regimes (Rollins 2009). They can also be readily associated with wildlife species where habitat requirements are developed by the U.S. Gap Analysis Program (Gergely and McKerrow 2013). Use of vegetation classification hierarchy, systematic treatment of the "natural-to-cultural" land cover continuum, handling of field observations, and spatial modelling with remotely sensed data all contribute to advanced LANDFIRE map products.

Our second case study describes continental-scale mapping that encompasses temperate and tropical latitudes of North America and all of South America. Mapping methods such as those used by LANDFIRE were adapted to the hemisphere to provide a more thematically detailed view than had been previously attained. The intent of this effort was to support conservation status assessment of ecosystem types occurring within and across national borders (Comer et al. 2020).

Both case studies used a supervised modeling approach which include *a priori* classification of vegetation types as the basis for mapping (Cihlar 2000, De Cáceres and Wiser 2012). That is, each began with a set of classification concepts where vegetation types are known and described. Georeferenced locations for these types are used by machine learning algorithms to train a predictive model and generate a map of their distribution (Muchoney et al. 2000). This approach was well suited to these efforts because development of a predictive distribution of described natural ecosystem types prior to intensive human intervention was needed for both case studies.

Table 1. U.S. National Vegetation Classification Hierarchy, including example classification units. The number of natural types documented within each hierarchical level from the conterminous United States (as of March 2021).

Level No.	Level Name	Defining Characteristics	No. Types	Example
1	Class	Life Form Physiognomy	6	Grassland & Shrubland
2	Subclass	Global Physiognomy	13	Temperate & Boreal Grassland & Shrubland
3	Formation	Global Physiognomy	36	Temperate Grassland & Shrubland
4	Division	Continental Floristics	50	Great Plains Grassland & Shrubland
5	Macrogroup	Subcontinental Floristics	143	Great Plains Tallgrass Prairie
6	Group	Regional Floristics	327	Northern Great Plains Tallgrass Prairie
7	Alliance	Subregional Floristics	1,174	*Schizachyrium scoparium - Bouteloua curtipendula* Northern Grassland
8	Association	Local Floristics	6,108	*Schizachyrium scoparium - Bouteloua curtipendula - Hesperostipa spartea -* (*Pascopyrum smithii*) Grassland

Case study 1 – Conterminous U.S. (LANDFIRE)

Many terrestrial ecosystems across temperate North America support natural wildfire regimes of varying frequency and intensity. The multi-agency LANDFIRE effort was established in 2001 to produce a series of moderate-to-high resolution map products, along with state-and-transition models, to characterize expected and actual vegetation condition with regards to natural disturbances like wildfire. All products of the effort are intended to facilitate strategic decisions by both wildfire and wildlife habitat managers.

Beginning with a vegetation-based classification standard, conceptual and quantitative state-and-transition models describe expected succession and disturbance pathways, as well as characteristic fuels, for a given vegetation type. Spatial models, called biophysical settings, aim to depict the likely historical location of each type, given natural disturbance regimes. These models were based on the terrestrial ecological systems classification developed by NatureServe (Comer et al. 2003). Then, current land cover map products aim to depict the location of each natural vegetation type, cultural land cover, vegetation canopy closure, canopy height, and recent disturbances. The ecological systems classification was developed in the early 2000s in part to address deficiencies in the U.S. National Vegetation Classification Standard, as it existed at the time (Federal Geographic Data Committee [FGDC] 1997). Specifically, the ecological systems classification established classification units that integrated geophysical characterstics with natural disturbance regimes to describe recurring assemblages of plant communities (Comer et al. 2003, Comer and Schulz 2007) with 638 units currently described for the conterminous USA. The national application of that classification led to substantial revisions to the U.S. National Vegetation Classification (USNVC), following the newly established IVC framework (Faber-Langendoen et al. 2014). This included additional hierarchical levels, making the USNVC more usable for mapping applications. In addition to using the ecological systems classification, the second major national remap effort by LANDFIRE adopted use of the USNVC Group-level concepts (see below) for mapping existing vegetation and land cover. In this case study, we will focus on this aspect of the LANDFIRE products.

Mapping units

The LANDFIRE map legend for existing vegetation and land cover encompasses the continuum from natural to ruderal and cultural vegetation types. The hierarchical structure of the USNVC includes broad units at upper levels defined by vegetation physiognomy, followed by progressively narrow units at lower levels defined by vegetation floristic composition (Federal Geographic Data Committee [FGDC] 2008). The full spectrum, from "natural" to "cultural" vegetation types are encompassed by the USNVC, but here we will reference use of "natural" to "ruderal" units. Table 1 provides an example of the USNVC hierarchy, with defining characteristics. Here, tallgrass prairie types have been well described at all levels of the hierarchy down to the association level, where multiple dominant and diagnostic species are used to define a given type. Over 6,000 associations describe natural vegetation types within the conterminous United States (CONUS). While this level of thematic detail is not currently feasible to map beyond relatively local scales, Group and Alliance levels are increasingly feasible to target in regional and national map legends. Within the CONUS, LANDFIRE mapped nearly 300 natural USNVC Group concepts. Descriptions of this classification hierarchy and these units may be found at the USNVC website (http://USNVC.org/) and NatureServe Explorer (https://explorer.natureserve.org/).

While the USNVC Group level provided a useful classification of natural vegetation units for the map legend, additional map legend categories were used to provide robust map product. First, the revised USNVC includes "ruderal" units that are defined as including plant assemblages with no natural analog. These commonly result from prior land conversion and subsequent abandonment, so they encompass what are often referred to as "old fields" and secondary forests where exotic species and/or native species are present in abundances not found where prior human influence is less discernable. Several ruderal vegetation units, approximating the Group level, were documented for use in LANDFIRE map products.

Second, a series of map class modifiers were used to facilitate mapping structural variants within each USNVC Group-based map class informed by the using the National Land Cover Database (NLCD; Homer et al. 2015). For example, where "evergreen" "deciduous" and "mixed" variants of a given forest type were discernable, they were mapped separately. Additionally, where "forest" vs. "shrubland" or "herbaceous" structural stages in forest succession occurred in discernable pattern, they were also differentiated in the map legend.

Training samples

Modern mapping methods include use of georeferenced sample locations – each labeled to the intended map units they represent – to train models that will combine predictor layers to generate a vegetation map. Due the the very large number of georeferenced samples needed for national land cover mapping at thematic levels like the USNVC Group, LANDFIRE produced algorithmic tools called "autokeys" for processing vegetation sample plot data for subsequent modeling and mapping. The autokey algorithm scans the content of each sample plot to detect species presence and abundance as well as structural categories to determine to which map legend class the sample belongs. It then applies the appropriate label for use in subsequent modeling steps. Autokeys were designed and implemented within regions determined by clustering ecologically similar ecoregions modifed from US Forest Service (Cleland et

al. 2007) and US Environmental Protection Agency (EPA) (Omernik and Griffith 2014) sources. In the CONUS project area 278 USNVC Groups were processed for use in the LANDFIRE legend. Each autokey pertains to one of 17 regions, including the southern tip of Florida which was treated along with adjacent Caribbean islands.

Expert ecologists reviewed and hand labeled nearly 18,000 samples to assess autokey performance. The total number of samples labeled by autokeys varied by region, from a high of 80,148 in the Rocky Mountains to 3,517 in the North Coast region. For most regions, the proportion of plots used in assessment that were reviewed by experts was 4% to 8%. Validation statistics for each map legend category were used throughout the development and final evaluation of each autokey. The overall validation statistic is a useful measure of how well each autokey performed across all types. It is the number of matches between expert and autokey labels divided by the total number of expert plots × 100, and the overall validation statistic was calculated for each autokey. Overall agreement for the USNVC Group keys ranged from a high of 90% (Texas-Oklahoma Hill Prairie) to a low of 39.9% (Coastal Plain). In most cases, lower performance occurred where substantial proportions of the landscape are dominated by ruderal vegetation, and distinguishing among very similar vegetation types using sample plots becomes more challenging.

Although over 500,000 vegetation samples were labeled through autokeys, there were still hundreds of thousands of samples with insufficient quantitative information to run through the autokeys. These often included documented locations from local natural resource inventories where an existing classification was used to label the location without including vegetation composition and structure. A series of classification crosswalks were used to reconcile these differences and label samples to the intended unit on the LANDFIRE map legend. In the 2016 LANDFIRE map, over one million samples were processed either by autokeys or through expert labeling across the CONUS.

Modeling process and resulting map

A key factor in ecosystem modeling is determining the boundaries within which to build and apply models for existing vegetation types. Initial prototyping found that Omernik Level III Ecoregions (Omernik and Griffith 2014) provided a more ecologically meaningful framework for modeling existing vegetation types than map zones used in previous LANDFIRE mapping efforts (Picotte et al. 2019). The ecoregions were grouped to create 13 vegetation production units (Figure 1) across the CONUS that were of a manageable size for efficiently preparing satellite

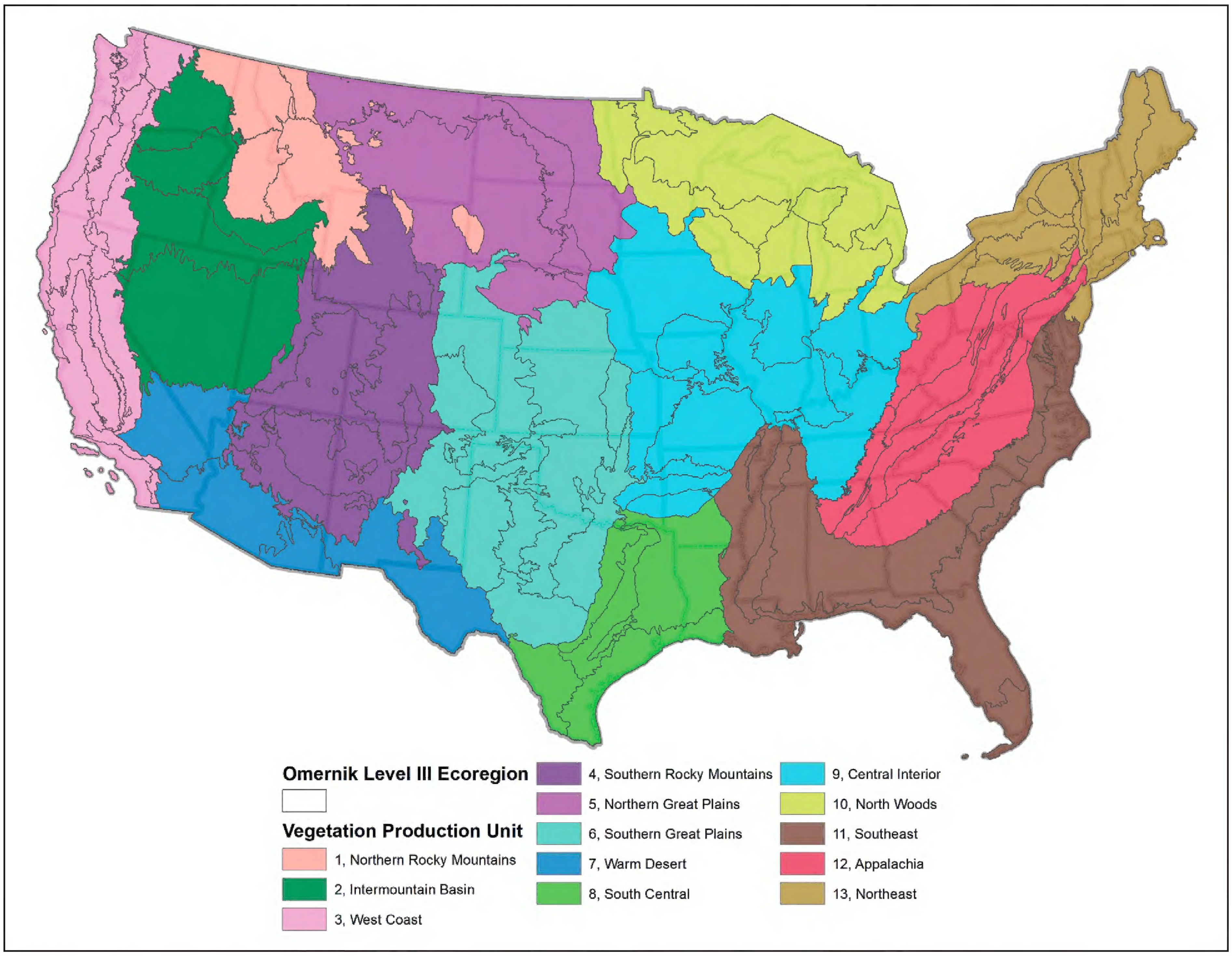

Figure 1. LANDFIRE Vegetation Production Units created by grouping Omernik Level III Ecoregions.

Table 2. Raster inputs to mapping vegetation types per 30 m pixel.

Dataset Name	Units of Analysis	Range of Values	Source (citation)
Elevation	Meter	-113 to 4415	3DEP DEM (USGS 2016b)
Aspect	Degrees	0 to 359°	3DEP DEM Derivative (USGS 2016b)
Percent Slope	%	0 to 85	3DEP DEM Derivative (USGS 2016b)
Topographic Position Index (300 & 2000)	Index	approx. 900–2,330, 400–3150	3DEP DEM Derivative (USGS 2016b)
Landsat Imagery (Seasonal) ca. 2016	Radiance	6 (0 to 255 per band)	Processed Landsat scenes courtesy of the U.S. Geological Survey
Tasseled Cap (Seasonal) ca. 2016	Index	3 (-8,000 to 24,000 per band)	Processed Landsat scenes courtesy of the U.S. Geological Survey
NDVI 5-year statistics (Min, Max, Median, Max-Median)	Index	-1.0 to 1.0	Processed Landsat scenes courtesy of the U.S. Geological Survey
Climate (Precipitation, Temperature)	Milimeters, Degree	1,390 to 65,534, 26,000 to 49,494	Gradients (Rollins and Frame 2006)
Soils (Percent Sand, Silt, Clay, Organic Matter, and pH)	%	0 to 100, 0 to 9	gSSURGO (USDA. 2016)

imagery and georeferenced sample inputs. These were similar to, but not identical to the ecoregion clusters used for design and implementation of autokeys.

The modeling process began by removing sample plots in recently disturbed areas collected prior to the disturbance using LANDFIRE annual disturbance products. Spectral outliers were identified by summing Landsat bands one through six for each class and sample plots, those plots greater than two standard deviations from the mean were removed. The resultant filtered plots were used to model lifeform and vegetation structure. The spectral test was performed separately for vegetation types prior to witholding samples for map validation.

Vegetation structure is important for fire behavior fuel models. Therefore, existing vegetation products were designed to nest by lifeform. For example, pixels identified as tree in the lifeform mask will be assigned a tree cover, height, and vegetation type. The lifeform modeling process began with an initial output using the filtered sample plots. The initial lifeform model output was improved through an iterative process by adding expert-labeled training samples based on desktop review of aerial photos to correct obvious mapping errors.

Using USNVC Group concepts as a guide, sample plots were separated into three lifeforms: tree, shrub, and herbaceous vegetation types, as well as barren or sparsely vegetated types (<10% total cover). Plots were further separated into wetland vs non-wetland categories, and alpine vs non-alpine categories where they existed. Classification tree models were generated with the See5 algorithm using raster predictor variables (Table 2). Models were generated within individual ecoregions to produce categorical outputs for each lifeform layer. The layers were then combined using the lifeform mask, wetland mask, barren/sparse mask, and alpine mask created with a separate modeling process to restrict the mapping of certain vegetation in appropriate locations where applicable. After modeling, vegetation types that comprised a mix of evergreen and deciduous dominant/co-dominant species to varying degrees were separated using the NLCD (Homer et al. 2015) categories to further refine the map and aid in assessing fire behavior fuel models. An overview of the modeling process is shown in Figure 2.

The draft thematic map was edited using rulesets based on geography or topography, or manual pixel reclassification with hand-drawn polygons based on expert opinion and review. Draft maps were also revised by removing problematic plots identified during the modeling process, reclassifying plots to a better fit, or adding sample plots based on expert opinion to correct modeling errors and improve mapping of problematic classes. Draft maps were reviewed by regional experts with NatureServe and staff from state agencies. Wildland Urban Interface maps produced by the Forest Service (Radeloff et al. 2017) were used in combination with rulesets based on existing vegetation to identify ruderal vegetation in proximity to developed areas. Recently disturbed areas (within previous 10-years) were identified using LANDFIRE disturbance products to appropriately label transitional vegetation. For example, it was more appropriate to label regenerating clearcuts in the Pacific Northwest as recently disturbed herbaceous cover rather than native montane grassland.

Vegetation percent cover and height training samples were derived from field estimates of vegetation cover and height. In addition, tree canopy percent cover estimates were calculated from the percentage of Light Detection and Ranging (LiDAR) point cloud above 3 m and tree height estimates were derived from the 90th percentile of LiDAR returns. Sample plots were separated into three lifeforms: tree, shrub, and herbaceous cover and height. Regression tree models predicting percent cover and height of dominant vegetation were generated with the Cubist algorithm using the following predictor layers identified in Table 2: Seasonal Landsat imagery, tasseled cap, and NDVI 5-year statistics. Models were generated within a vegetation production unit to predict continuous outputs for each lifeform layer. The layers were then combined using the lifeform mask. Recently disturbed areas (within previous 3 years) did not model well because the satellite image composites spanned multiple years and comprised a mixture of pre and post-disturbance pixels. Disturbance severity and timing rulesets based on LANDFIRE annual disturbance products were used to assign lifeform and estimated cover.

Several masks were developed to identify open water, barren land, sparse vegetation, developed, and agricultural lands. Open water was identified using custom model-

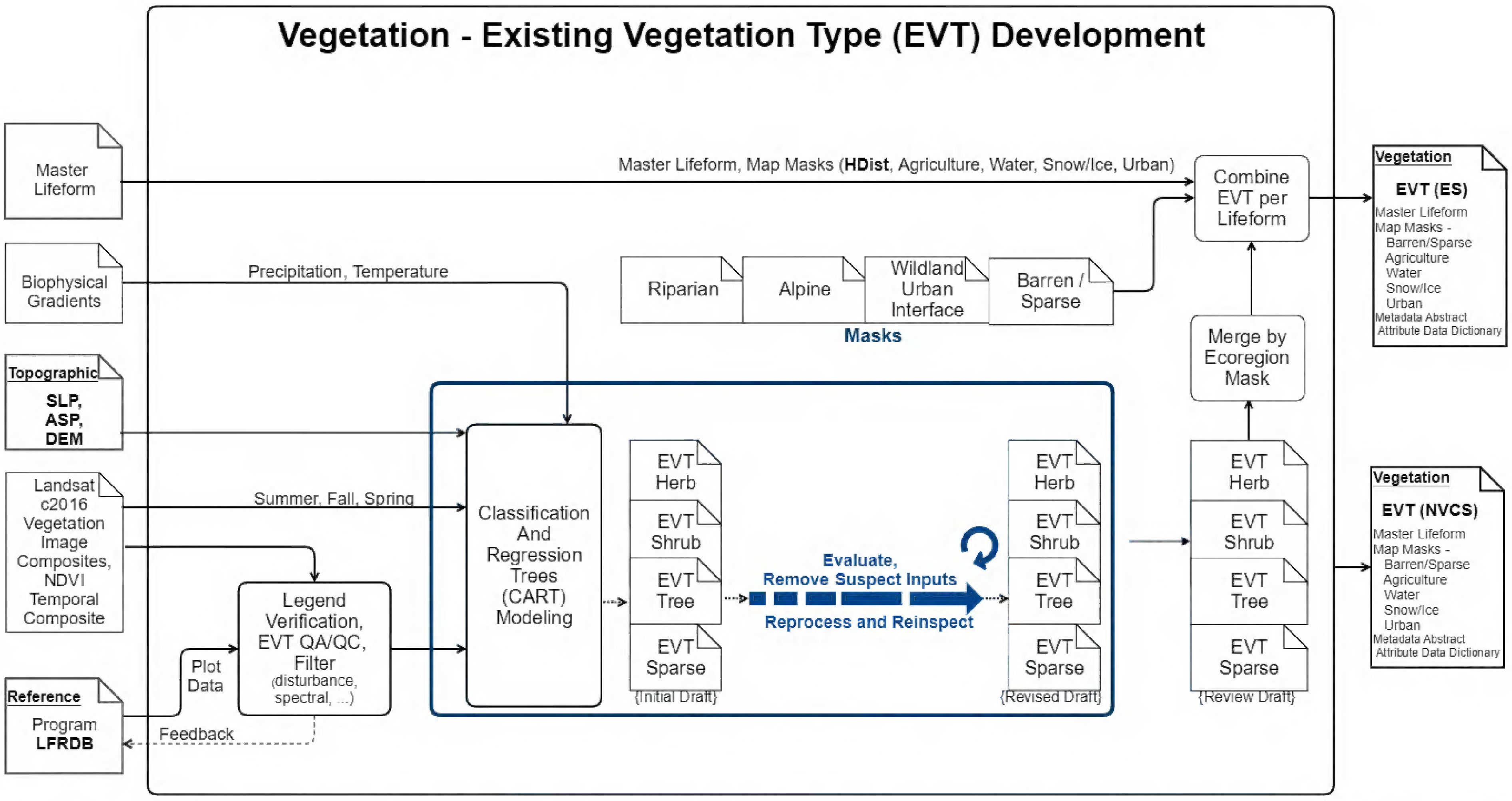

Figure 2. LANDFIRE Existing Vegetation Type (EVT) Development flow chart; additional details can be found in the metadata abstract available online at LANDFIRE Remap 2016 National Vegetation Classification (NVC) CONUS (usgs.gov)

ing methods based on the Landsat Level 3 Dynamic Surface Water Extent (DSWE) Science Product from the U.S. Geological Survey. Fragmented segments along streams and rivers were connected using National Hydrography Dataset (USGS 2016a) flowlines to form a continuous network. Barren land (0% vegetated cover) and sparse vegetation (<9% total cover) were identified using NDVI 5-year median thresholds and calibrated by location based on known barren and sparsely vegetated areas. Developed land and snow/ice were identified using NLCD. Ruderal vegetation classes were modeled within NLCD Class 21 (Developed – Open Space) in order to assign the appropriate fire behavior fuel models. Agricultural land was identified using the Cropland Data Layer (USDA 2015) and summarized by Common Land Unit polygons (USDA 2006) using a zonal majority.

The mapping process for existing vegetation based on the USNVC Group concepts generated a 30 m pixel resolution map raster with 499 natural, ruderal, and cultural map classes (Figure 3). The majority of these map classes are reflected in the more than 300 USNVC Groups. In addition to these, map class modifiers distinguish structural variants within each natural USNVC Group. Then a series of map classes reflect USNVC Groups for ruderal vegetation plus cultural land cover derived from other sources as described above.

Evaluation of LANDFIRE map

The LANDFIRE Program implemented a map assessment approach that utilized existing information because no resources were available to collect additional data. The goal was to provide assessment results in tandem with data delivery. The assessment sampling strategy for the LANDFIRE 2016 Remap randomly withdrew 10% of the available plots for each of the terrestrial ecological systems classification developed by NatureServe (Comer et al. 2003) within a vegetation production unit, with two caveats. First, no more than 300 plots were withdrawn for any ecological system type. Second, no plots were withdrawn for an ecological system type within an individual production unit if less than 30 plots would be available for the assessment in an effort to maximize training data for modeling categories with few plots. This same set of withdrawn plots was used for the USNVC assessment, although the plot labels were assigned using independent methods and therefore did not always represent conceptually equivalent types. Withdrawn plots were never used in the modeling process so they represent an independent assessment sample.

Confusion tables were created for each of the six LANDFIRE GeoArea delivery packages across CONUS by cross-tabulating the autokey USNVC Group assignment for each assessment plot against the LANDFIRE USNVC Group assignment for map pixels at the plot location. Category agreement focused tables were then generated from each GeoArea contingency table. No stratification, spatial buffering, or category weighting was used (Table 3). A summary report was also generated for each GeoArea to provide an initial indication of map performance (available from https://www.landfire.gov/remapevt_assessment.php).

The assessment sample was based on plots previously available to the program so the sample size and distribution reflected the overall plot numbers and categorical

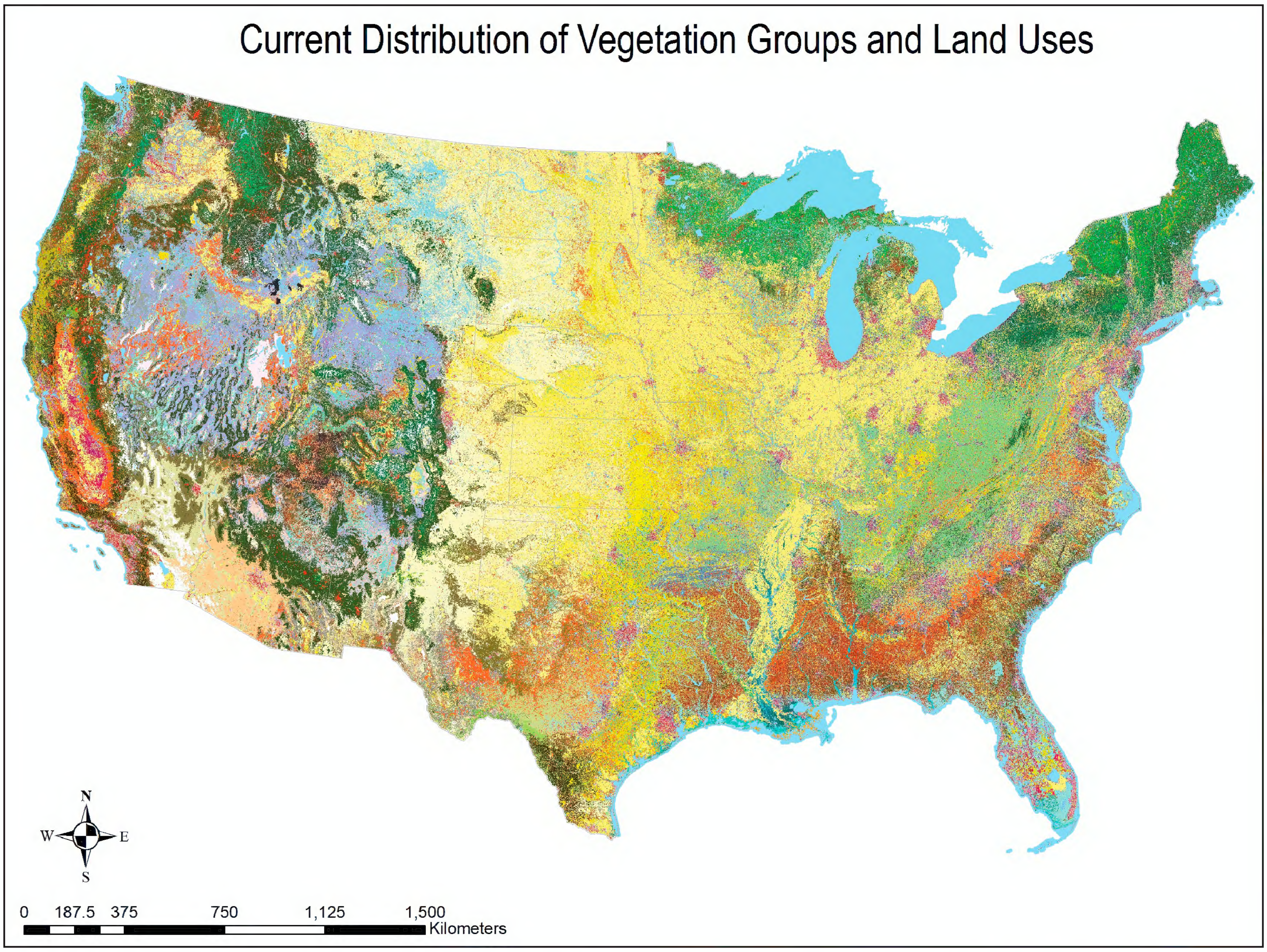

Figure 3. USNVC Vegetation Groups, modified to include structural variants, plus land uses, encompassing 499 natural, ruderal, and cultural map classes).

Table 3. Assessment results for USNVC Groups with sufficient samples (n > 30) by GeoArea.

GeoArea	No. USNVC Groups with assessment plots	No. USNVC Groups with >30 assessment plots	Proportion of USNVC Groups with >30 assessment plots with >70% agreement between map and plot designation	Proportion of USNVC Groups with >30 assessment plots with >50% agreement between map and plot designation
Northwest	132	38	13%	34%
North Central	95	24	8%	29%
Northeast	139	48	10%	40%
Southwest	173	61	15%	41%
South Central	122	17	35%	59%
Southeast	50	19	5%	42%

distribution present in any GeoArea. Across the continent the assessment sample was not sufficient for most of the mapped categories. While the agreement results were not high, there was variation in the results across GeoAreas and across categories within each GeoArea. No consistent error patterns were identified, although there is some indication that forest types tend to have lower error rates than shrub and herbaceous types. The opportunities for comparing category error rates across GeoAreas are limited by the sample sizes. For example, Southern Rocky Mountain Ponderosa Pine was mapped in both the Southwest and Northwest GeoAreas but the limited extent in the Northwest GeoArea resulted in too small an assessment set for this GeoArea. Results were not specifically linked to the number of categories assessed. For example, while the Southeast GeoArea had the lowest number of assessed categories and the lowest percentage of assessed categories with more than 70% agreement, the North Central had the lowest percentage of assessed categories with more than 50% agreement. Readers should note that these are absolute errors. If the plot assignment did not match the map assignment exactly it was designated as an error, so errors between floristically similar groups are counted the same as errors between floristically dissimilar groups. Users can review results for USNVC Groups of specific interest to fully understand the results of the assessment analysis.

To understand the results and ramifications for the US-NVC Group-based map, a small portion of the Category Agreement Table for the Northwest GeoArea is presented in Table 4. One row represents higher agreement, one row

Table 4. Portions of the Northwest GeoArea Category Agreement Table for USNVC Group.

USNVC Name	Row Total (pixels)	Row Agreement	Primary Within Row Mismatch	Secondary Within Row Mismatch	Tertiary Within Row Mismatch
Columbia Plateau Western Juniper Woodland & Savanna	89	75.28%	6480 Columbia Plateau Western Juniper Shrubland; 4 Incorrect Pixels	6288 Intermountain Mountain Big Sagebrush Shrubland & Steppe; 3 Incorrect Pixels	6145 Central Rocky Mountain Lower Montane Foothill & Valley Grassland; 3 Incorrect Pixels
Intermountain Basins Dry Tall Big Sagebrush Shrubland & Steppe	740	56.49%	6285 Intermountain Low & Black Sagebrush Shrubland & Steppe; 58 Incorrect Pixels	6287 Intermountain Mesic Tall Sagebrush Shrubland & Steppe; 36 Incorrect Pixels	6070 Rocky Mountain Subalpine Dry-Mesic Spruce - Fir Forest & Woodland; 28 Incorrect Pixels
Vancouverian & Rocky Mountain Montane Wet Meadow & Marsh	35	22.86%	6239 Western Montane-Subalpine Riparian & Seep Shrubland; 4 Incorrect Pixels	6070 Rocky Mountain Subalpine Dry-Mesic Spruce - Fir Forest & Woodland; 3 Incorrect Pixels	6330 Northern Rocky Mountain Lowland & Foothill Riparian Forest; 2 Incorrect Pixels

represents moderate agreement and one row represents low agreement. The key information presented in this table is what type of errors were made, not just overall agreement.

For example, the most prevalent misclassification for Intermountain Basins Dry Tall Sagebrush Shrubland & Steppe was Intermountain Low & Black Sagebrush Shrubland & Steppe, followed by Mesic Tall Sagebrush Shrubland & Steppe. These types can occur immediately adjacent to each other across the western landscapes where they are found, and share substantial floristic composition, while the Dry-Mesic Spruce - Fir Forest & Woodland is much less similar so those errors may be more substantial depending on the application. This type of variation in agreement results was common across the GeoAreas, so users can review results for USNVC Groups of specific interest to fully understand the results of the assessment analysis.

Map users should also note that, in addition to the issues with sample size and distribution, these results do not indicate the scope of misclassifications, e.g., how much area within a GeoArea had agreement greater than or less than 50% or 70%.

Case study 2 - Continental Americas (NatureServe)

Accelerating landscape change threatens biodiversity worldwide, so documented trends in the extent of ecosystems provide a foundation that can be used for conservation action. However, a comprehensive ecosystem classification of sufficient thematic detail to support these types of analyses has been lacking across the Americas. While a number of ecosystem classification maps exist at regional (Sano et al. 2010), continental (Stone et al. 1994, Eva et al. 2004), and global (Sayre et al. 2020) extents, nearly all utilize thematic classifications with a limited number of land unit descriptors that do not differentiate floristic composition among types.

This second case study includes a project area of approximately 32.6 million km^2 or nearly 22% of the global land surface, excluding the Boreal and Arctic regions of North America. The aim was to produce both "potential" and "current" distribution maps for major terrestrial ecosystem types that would be suitable for continental-scale assessment and planning, and also include units suitable for on-the-ground conservation action. The "potential distribution" includes biophysical conditions where each type might occur today had there not been any prior intensive human intervention. "Current distribution" then accounts for those areas of intensive intervention and conversion, as of approximately 2010. For this effort an effective minimum map unit size, or mapped pixel resolution, ranged from 270 m to 450 m.

Mapping across the hemisphere brings challenges of working with a high diversity of vegetation types and uneven availability of modeling inputs. Different modeling approaches and lower levels of thematic and spatial resolution in map products may be useful. Building from experience in the United States, we developed new spatial models of potential distributions of vegetation Macrogroups as defined by the International Vegetation Classification or IVC (Faber-Langendoen et al. 2014) (Table 1). We then combined current land use classes, derived from globally-available land use maps, with potential distribution maps of natural types to estimate their current extent. Similar modeling methods previously applied in Africa (Sayre et al. 2013) were adapted for this effort in the Americas. Analytically, RandomForest (Gislason et al. 2006) classification and regression trees (CART) (Breiman et al. 1984) were used to identify relationships of predictor layers for combinations of map surfaces relative to the location of georeferenced samples for each target class from the desired map legend (Hansen et al. 1996, De'ath and Fabricus 2000). A combination of ArcGIS (10.1), ERDAS Imagine, and the data mining tool See 5 (Rulequest Research 2012) was used to develop models representing vegetation type distributions.

Mapping inputs

Table 5 provides a summary of map inputs, including existing map sources for potential distribution modeling. Here we emphasize project components outside the USA. Existing national and regional maps, along with georeferenced field sample data for vegetation types, were all reconciled thematically to the IVC and NatureServe ecological systems classifications (Comer et al. 2003, Josse et al. 2003). Again, given the intent to map potential distribution of "natural" vegetation, only these types were sampled from existing sources. Given limitations of available field samples, randomized samples were also gathered

Table 5. Map sources of sample points for model development – from institutions and publications - with emphasis on Latin America and Caribbean (MMU = minimum map unit).

Mapping Region	Map Sources	Source MMU	Sample Points
Caribbean	Borhidi (1991), The Nature Conservancy (TNC)	1 ha	80,539
Mexico	Mexican National Institute of Statistics and Geography (INEGI), TNC, ProNatura-Yucatan	5 ha	41,731
MesoAmerica	TNC, ProNatura-Yucatan	5 ha	56,372
South America	TNC, NatureServe, World Wildlife Fund	1000 ha	416,309

from existing local maps in order to provide a robust and spatially balanced representation of each target map class where there was an acceptable level of confidence in the map source. Here we define "acceptable" as being judged sufficiently reliable by project ecologists experienced in the region and familiar with each map source.

Next, we screened the patch sizes of a given type, patches > 10 km^2 in area provided the pool of source areas for sample selection. Selection of the 10 km^2 is again an expert judgment, having evaluated existing maps and concluded that sampling from types depicted in smaller areas risked introducing substantial error. We acknowledge that this risks exclusion of naturally rare ecosystem types, but we judged this risk was warranted given the quality of existing map information for this purpose. This pool of map polygons encompasses 95% of natural landscapes. Stratified random sample selection was weighted by continent-wide area of each type using the $\log_{10}$ (area)*100, providing a sample total weighted towards types of lesser area. A total of 595,951 georeferenced samples were generated for the Americas, with an additional 70,380 held aside for map validation.

Explanatory variables, represented as map surfaces, included a series of biophysical factors, such as bioclimate, landform, slope, and aspect, as well as surface flow accumulation (Table 6).

Bioclimates, as modeled by Metzger et al. (2013), reflect the categorization of temperature and precipitation regime to globally-available remotely sensed data, resulting in a total of 125 unique bioclimates at 1 km^2 spatial resolution. Geophysical map surfaces were developed using 90 m × 90 m digital elevation data (Jarvis et al. 2008). Slope and aspect were measured in terms of degrees. The methodology for the landform class derivation used a variable moving window to assess relative relief and followed other regional scale approaches to model macro-landforms (Dikau et al. 1991, True et al. 2000). Landforms as discrete units were derived from Weiss (2001), who used the 90 m continental digital elevation to assign pixels into one of the following regional physiographic types: "canyons, deeply incised streams," "midslope drainages, shallow valleys," "upland drainages, headwaters," "u-shaped valleys plains," "open slopes," "upper slopes, mesas," "local ridges, hills in valleys," "midslope ridges, small hills in plains," and "mountain tops, high ridges." Land surface flow accumulation was derived from existing continental data (Lehner et al. 2006), based on a 90 m × 90 m hydrologically conditioned digital elevation. This data set specifically aims to use topographic surfaces to indicate stream flow direction and flow accumulation for use in stream, wetland, and riparian ecosystem analysis. EarthSat NaturalVue (1998–2002) multi-year composites of a red-green-blue translation of 6 bands from Landsat 5–7 at 150 m pixel resolution served as the only spectral inputs to the model. These data served to differentiate the locations of natural types tending to occur in proximity, and therefore, similar to geophysical settings. An overview of the hierarchical modeling process to map potential distibrutions of IVC Macrogroups is shown in Figure 4.

IVC vegetation hierarchy modeling

We used a sequential mapping process where maps derived for multiple broader levels of the IVC classification hierarchy were then used as input to modeling distributions of types defined at lower hierarchical levels. In this application, the first thematic level for inductive modeling was the IVC Division (Level 4 from Table 1). For example, in South America the 80 types may be viewed as continental expressions of vegetation formations; with vegetation responding most directly to global climate patterns. On average, 1,032 (Min = 53, Max = 5,724) samples per map class were used to generate the South America portion of this map. No satellite imagery was used in the development of the IVC division-level

Table 6. Inputs to mapping vegetation types (as needed for modeling, each layer was rescaled to summarize variable per 90 m pixel).

Dataset Name	Data Type	Range of Values	Spatial Resolution	Institutional Source (or citation)
Climate	raster	125	1 km^2	Metzger et al. (2013)
Slope	raster	89	90 m × 90 m	NatureServe, from SRTM digital elevation
Aspect	raster	1–360°	90 m × 90 m	NatureServe, from SRTM digital elevation
Landform	raster	11	90 m × 90 m	NatureServe, from SRTM digital elevation
Lithology	raster	9–40	450 m × 450 m	Sayre et al. (2008), INEGI (*Mexico*), USGS (*Caribbean*)
Soils	raster	259	90 m × 90 m	CanVec (Natural Resources Canada)
Surface Flow Accumulation	raster	156	90 m × 90 m	HydroSHEDS (Lehner et al. 2006)
EarthSat NatureVue Imagery	raster	3 (0–255 per band)	150 m × 150 m	ESRI
Map Samples	raster	683,119	90 m × 90 m	LANDFIRE, Josse et al. (2007, 2009), Borhidi (1991), and others
Hexagon Grid	vector	320,561	96 km^2	NatureServe, DGGRID, Sahr (2013)

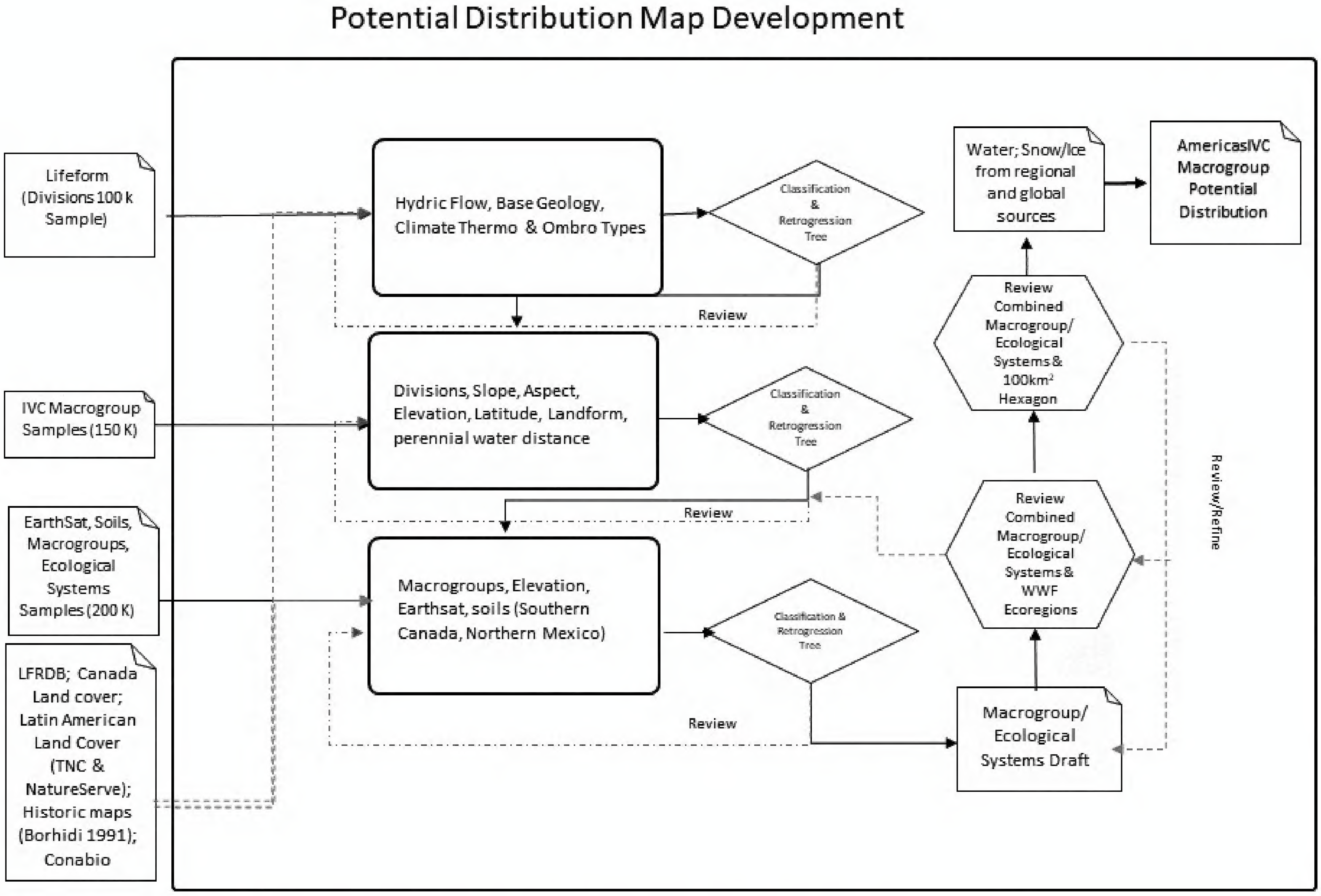

Figure 4. Potential Distribution Map Development flow chart for Case Study 2.

map output, the EarthSat NaturalVue (1998–2002) imagery was used for both MacroGroup and Ecological Systems level of classfication. Macrogroups were subsequently modeled using an average of 1,234 (Min = 33, Max = 3,433) samples per map class. Both the IVC division map output and EarthSat NaturalVue (1998–2002) imagery were used as map inputs for the macrogroup model. Terrestrial ecological systems, being most numerous and most finely differentiated among the classification units used in this effort, were modeled using the macrogroup map as an additional model input. Once completed, the modeled terrestrial ecological systems layer is the finest thematic scale achievable using this technique. Because these units could be conceptually nested into IVC macrogroup concepts, the "bottom-up" aggregation of maps depicting these units is expected to provide the most reliable map of macrogroups. This aggregation was then reviewed and edited to finalize the distribution of each IVC macrogroup (Figure 5). Numbers of map classes by region and classification level are listed in Table 7.

Table 7. Numbers of mapped classification units by region and level of ecological classification (including types with regionally overlapping distributions).

Region	Number of IVC Divisions	Number of IVC Macrogroups
Caribbean	4	14
Mexico	32	55
Central America	20	33
South America	80	190

Map editing and refinement

Over-prediction of more common (over rare, or low sample size) land cover types is a common source of error in CART-based inductive modeling of land cover (Weiss 1995, Lowry et al. 2007). That is, more common land cover classes can be over-mapped at the expense of less common classes. This could be anticipated in this particular application where there is high similarity in predictor variable combinations for vegetation types that are naturally adjacent. In these instances, predicted distributions may be skewed in favor of some over other types in portions of their range of co-occurrence. Because of this, some form of expert-based review and map refinement is unavoidable. We assumed that over-prediction would be concentrated in regional landscapes with extensive land use history and only fragmentary remnants of natural vegetation types. However, because the generalized distribution of each terrestrial ecological system type had been previously documented by country and World Wildlife Fund ecoregion (Josse et al. 2003), this knowledge was used in expert type-by-type review and refinement. Draft model outputs were attributed as extent measures per WWF ecoregion. These distributions were compared against known ecoregion distributions to identify likely error. Types found to be in error had their pixel distributions recoded to most-likely correct types for each WWF ecoregion. In turn, a second phase expert review and refinement followed the procedure used in ecoregions but was applied

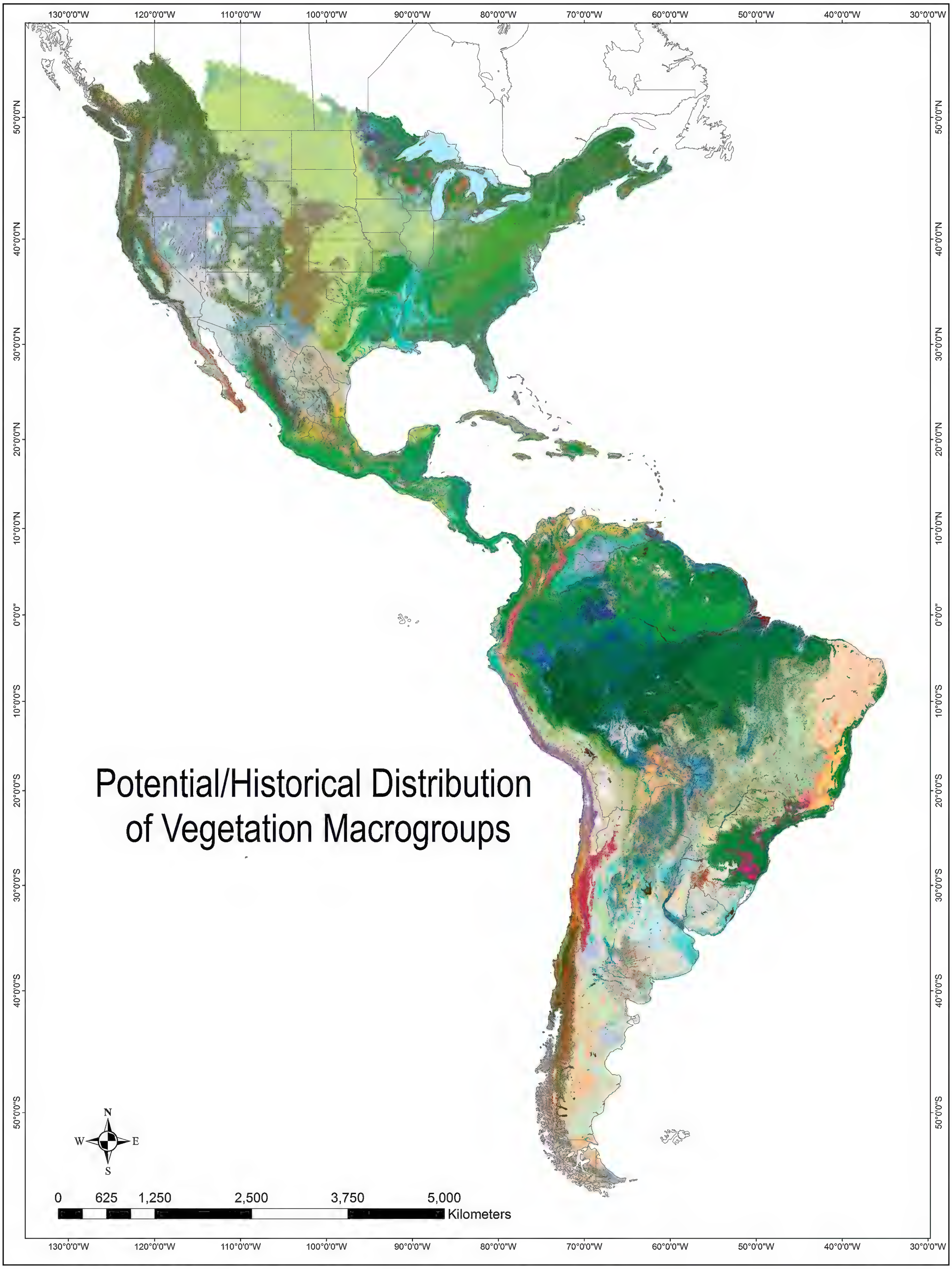

Figure 5. Potential/historical distribution of IVC Macrogroups (at 270 m pixel resolution and 315 natural map classes).

to each type using a common grid of 100 km² hexagons (Sahr 2013). Again, with each type attributed to the hexagon grid, type-by-type review led to final recoding of pixels to most-likely correct types. Final map products were produced at 90 m and 270 m pixel resolutions with resampling to the unified pixel size accomplished using bilinear interpolation technique suitable for continuous data.

Map validation

As noted above, during initial sample data collection from map sources, georeferenced samples of each vegetation type were gathered and set aside for use in map validation. These samples were gathered for types that had existing polygons in regional/local source maps > 10 km^2 in size in South America and > 5 hectares for temperate and tropical North America. Of the 315 Macrogroup map classes in North and South America, 284 had sufficient samples to be quantitatively assessed.

Once map edits were finalized for the 90 m products, validation samples were used to score the degree of agreement between samples and map classes for each map class at three spatial scales. Circular buffers around each sample encompassed 1-km^2 (within 6 pixels of center) and 5-km^2 (within 28 pixels of center). A point sample was defined from the centroid of each pixel of the 6 × 6 neighborhood of the 90 m product and is equivalent to the 270 m version of each map. Overlay of these samples on the final map product generated tabular summaries to determine whether or not the mapped class present matched the type labeled to each sample; i.e., the same types co-occur within the buffered area. While truly independent samples could not be acquired to evaluate a spatial model depicting "potential/historical" extent of these vegetation types, this technique provides one initial measure of map quality, and serves as a primary input to decisions regarding use of the map for type-by-type assessment. Thus, the percentage of agreement between validation samples and maps can indicate the degree of map reliability for use with a practical minimum map unit of 270 m vs. 1 km^2 vs. 5 km^2.

Table 8 provides a summary of validation statistics for potential distribution maps of macrogroups. Additional detail is found in Suppl. material 1. For the 315 mapped macrogroup types, per class numbers of samples for the 1-km^2 validation sample area varied from a high of 972 to a low of 10. Summary of validation statistics indicate high (> 80%) to moderate (> 60%) map accuracy overall, and on a per map class basis at 270 m vs. 1 km^2 vs. 5 km^2 map resolutions. Using the most demanding "point" (or 270 m) validation sample area, 8 types scored at 90–100% agreement, 11 types scored 80–90% agreement, 16 types scored 70–80% agreement, 39 types scored 60–70% agreement and 52 types scored 50–60% agreement. A total of 158 types (56% of all assessed map classes) scored below 50% agreement.

The inclusion of the 1 km^2 was limited to the North American portion of the map product for two reasons. First, the sample sizes available for CONUS in North America was substantially higher than that for adjacent countries. Secondly, the inclusion of the 1 km^2 allowed the examination of the gradient of model performance over a spatial gradient of neighborhoods. Using the 1-km^2 validation sample area in North America only, 44 types scored at 90–100% agreement, 32 types scored 80–90% agreement, 14 types scored 70–80% agreement, 8 types scored 60–70% agreement, and 9 types scored 50–60% agreement. A total of 12 types (10% of all assessed map classes) scored < 50% agreement. For 1 km^2 samples, the total sample agreement was 85% and the median level of map class agreement for the types assessed was 88%.

Using the 5 km^2 validation sample area, 160 types scored at 90–100% agreement, 50 types scored 80–90% agreement, 23 types scored 70–80% agreement, 20 types scored 60–70% agreement and 15 types scored 50–60% agreement. A total of 16 types (6% of all assessed map classes) scored below 50% agreement. For 5-km^2 samples, the total sample agreement was 85% and the median level of map class agreement was 92%.

These results indicate that map reliability is limited on a per pixel basis (at 270 m pixels), but within relatively small clusters of adjacent pixels, the reliability of the map increases for most map classes.

Conclusions

There is scientific value in documenting the location and trends in the extent and condition of ecosystem types to inform public policy and conservation action. These two case studies illustrate what can be accomplished with the systematic application of robust, hierarchically-structured vegetation-based classification and machine learning tools that utilize georeferenced sample locations and robust predictor maps.

The USA LANDFIRE case study illustrates where a deep history of vegetation-based classification and investments in key inputs to mapping (e.g., georeferenced samples, remote sensing data, sophisticated algorithms) can come together to generate standard map products covering more than 9.8 million km^2 of U.S. land that are unsurpassed, in terms of spatial and thematic resolution, anywhere in the world. That being said, it also remains clear that mapping at thematic resolutions of the USNVC Group and finer require very large and spatially balanced inputs of georeferenced samples, and even with the extensive prior investments, these remain a key limitation affecting the quality of map outputs. While one can reasonably say that "we know enough" about vegetation types at

Table 8. Summary validation statistics for 284 (315 total mapped) assessed macrogroups in North and South America.

Validation Sample Resolution	No. with 90–100% Agreement	No. with 80–90% Agreement	No. with 70–80% Agreement	No. with 60–70% Agreement	No. with 50–60% Agreement	No. with <50% Agreement
270 m (point) *n = 284*	8	11	16	39	52	158
*1 km^2 *n = 131*	44	32	14	8	9	12
5 km^2 *n = 284*	160	50	23	20	15	16

* North America only.

"mid" scales of the classification hierarchy (e.g., the USNVC Group), sufficient numbers of georeferenced samples that depict the full spectrum of those classification units is lacking across their range of distribution. Efforts such as LANDFIRE provide knowledge of where these gaps exist so that new data collection could maximize its effect on future map iterations.

The NatureServe effort for the Americas - encompassing 22% of the global land surface - demonstrates methods and outputs suitable for worldwide application at continental scales; albeit more challenging in parts of the globe with a more limited history of ecosystem classification and mapping, and more limited availability of predictor layers. Along with this mapping approach, the rich text, tabular, and map data set accompanying that study provide a foundation for deepened analysis and conservation action across the Americas. Continued collection of the input data used in the case studies could enable mapping at these spatial and thematic resolutions around the globe.

Data availability

Data associated with the LANDFIRE case study are available from www.landfire.gov; Data associated with the NatureServe case study are accessible from: https://transfer.natureserve.org/download/Longterm/Ecosystem_Americas/Maps/

Author contributions

P.J.C. was team member in both case studies, and led the manuscript preparation; J.C.H. developed and implemented map production of the second NatureServe case study; D.D. completed primary mapping tasks in the LANDFIRE case study, and provided manuscript text and review; J.S. coordinated with P.J.C. on classification-related efforts for LANDFIRE case study, and contributed text for the manuscript. All authors critically revised the manuscript.

Acknowledgments

Work by KBR was performed under USGS contract 140G0121D0001. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. The LANDFIRE portion was funded by U.S. Geological Survey (USGS) Earth Resources Observation and Science (EROS) Center.

References

Aycrigg JL, Beauvais G, Gotthardt T, Boykin K, Williams S, Lennartz S, Vierling KT, Martinuzzi S, Vierling LA (2010) Mapping species ranges and distribution models across the United States. Gap Analysis Bulletin 18: 12–20.

Borhidi AL (1991) Phytogeography and vegetation ecology of Cuba. Akademiai Kiado, Budapest, HU, 858 pp.

Breiman L, Friedman J, Stone CJ, Olshen RA (1984) Classification and regression trees. 1st ed. CRC Press, London UK, 368 pp.

Cihlar J (2000) Land cover mapping of large areas from satellites: Status and research priorities. International Journal of Remote Sensing 21(6–7): 1093–1114. https://doi.org/10.1080/014311600210092

Cleland DT, Freeouf JA, Keys Jr JE, Nowacki GJ, Carpenter C, McNab WH (2007) Ecological subregions: sections and subsections of the conterminous United States [1:3,500,000] [CD-ROM]. [Sloan, AM, cartog. Gen. Tech. Report WO-76]. Department of Agriculture, Forest Service, Washington, DC, US. https://doi.org/10.2737/WO-GTR-76D

Comer P, Schulz K (2007) Standardized ecological classification for meso-scale mapping in southwest United States. Rangeland Ecology and Management 60: 324–335. https://doi.org/10.2111/1551-5028(2007)60[324:SECFMM]2.0.CO;2

Comer P, Faber-Langendoen D, Evans R, Gawler S, Josse C, Kittel G, Menard S, Pyne M, Reid M, Teague J (2003) Ecological Systems of the United States: A working classification of U.S. terrestrial systems. NatureServe, Arlington, VA, US.

Comer PJ, Hak JC, Josse C, Smyth R (2020) Long-term loss in extent and current protection of terrestrial ecosystem diversity in the temperate and tropical America. PLoS ONE 15: e0234960. https://doi.org/10.1371/journal.pone.0234960

De Cáceres M, Wiser SK (2012) Towards consistency in vegetation classification. Journal of Vegetation Science 23: 387–393. https://doi.org/10.1111/j.1654-1103.2011.01354.x

De'ath G, Fabricius KE (2000) Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology 81: 3178–3192. https://doi.org/10.1890/0012-9658(2000)081[3178:CARTAP]2.0.CO;2

Dikau R, Brabb EE, Mark RM (1991) Landform classification of New Mexico by computer. U.S. Geological Survey [Open File Report 91–634]. Menlo Park CA, USA. https://doi.org/10.3133/ofr91634

Earth Satellite Corporation (2002) Economic and Social Research Institute. EarthSat NaturalVue Global Landsat Mosaic, 1998–2002. [GeoTiff]. ESRI. [Retrieved from] https://esri.com

Eva HD, Belward AS, De Miranda EE, Di Bella CM, Gond V, Huber O, Jones S, Sgrenzaroli M, Fritz S (2004) A land cover map of South America. Global Change Biology 10: 731–744. https://doi.org/10.1111/j.1529-8817.2003.00774.x

Faber-Langendoen D, Keeler-Wolf T, Meidinger D, Tart D, Josse C, Navarro G, Hoagland B, Ponomarenko S, Saucier J-P, … Comer P (2014) EcoVeg: a new approach to vegetation description and classification. Ecological Monographs 84: 533–561. https://doi.org/10.1890/13-2334.1

Faber-Langendoen D, Baldwin K, Peet RK, Meidinger D, Muldavin E, Keeler-Wolf T, Josse C (2018) The EcoVeg approach in the Americas: U.S., Canadian and international vegetation classifications. Phytocoenologia 48: 215–237. https://doi.org/10.1127/phyto/2017/0165

FGDC [Federal Geographic Data Committee] FGDC-STD-005-1997 (1997) National Vegetation Classification Standard, Version 1. Vegetation Subcommittee, U.S. Geological Survey, Reston, VA, US.

FGDC [Federal Geographic Data Committee] FGDC-STD-005-2008 (2008) National Vegetation Classification Standard, Version 2. Vegetation Subcommittee, U.S. Geological Survey, Reston, VA, US.

Gergely KJ, McKerrow A (2013) Species data – National inventory of range maps and distribution models: U.S. Geological Survey Fact Sheet 2013–3087, 1 pp. https://doi.org/10.3133/fs20133087

Gislason PO, Benediktsson JA, Sveinsson JR (2006) Random forests for land cover classification. Pattern Recognition Letters 27: 294–300. https://doi.org/10.1016/j.patrec.2005.08.011

Groves C, Game ET (2016) Conservation planning: informed decisions for a healthier planet. Roberts Publishers. Greenwood Village CO, USA, 580 pp.

Hansen M, Dubayah R, DeFries R (1996) Classification trees: an alternative to traditional land cover classifiers. International Journal of Remote Sensing 17: 1075–1081. https://doi.org/10.1080/01431169608949069

Homer CG, Dewitz JA, Yang L, Jin S, Danielson P, Xian G, Coulston J, Herold ND, Wickham J, Megown K (2015) Completion of the 2011 National Land Cover Database for the conterminous United States-Representing a decade of land cover change information. Photogrammetric Engineering and Remote Sensing 81: 345–354.

Jarvis A, Reuter HI, Nelson A, Guevara E (2008) Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database. https://bigdata.cgiar.org/srtm-90m-digital-elevation-database/

Josse C, Cuesta F, Navarro G, Barrena V, Cabrera E, Chacón-Moreno E, Ferreira W, Peralvo M, Saito J, Tovar A (2009) Mapa de ecosistemas de los Andes del norte y centrales. Bolivia, Colombia, Ecuador, Perú y Venezuela. Programa Regional ECOBONA-Intercooperation, CONDESAN-Proyecto páramo Andino, Programa BioAndes, EcoCiencia, NatureServe, IAvH, LTA-UNALM, ICAE-ULA, CDC-UNALM, RUMBOL SRL, Lima.

Josse C, Navarro G, Comer P, Evans R, Faber-Langendoen D, Fellows M, Kittel G, Menard S, Pyne M, Teague J (2003) Ecological Systems of Latin America and the Caribbean: A working classification of terrestrial systems. NatureServe, Arlington, VA, US.

Josse C, Navarro G, Encarnación F, Tovar A, Comer P, Ferreira W (2007) Sistemas ecológicos de la cuenca amazónica de Perú y Bolivia. Clasificación y mapeo. NatureServe, Arlington, VA, US.

Keith DA, Rodríguez JP, Rodríguez-Clark KM, Nicholson E, Aapala K, Alonso A, Asmussen M, Bachman S, Basset A, … Zambrano-Martínez S (2013) Scientific foundations for an IUCN Red List of Ecosystems. PLoS ONE 8: e62111. https://doi.org/10.1371/journal.pone.0062111

Lavrinenko IA (2020) Approaches of European ecologists to typology and mapping of habitats. Geobotanical mapping 2020: 51–77. https://doi.org/10.31111/geobotmap/2020.51

Lehner B, Verdin K, Jarvis A (2006) Technical Documentation Version 1.0. HydroSHEDS. World Wildlife Fund and partners. http://hydrosheds.cr.usgs.gov/index.php [accessed 26 August 2013]

Lowry J, Ramsey RD, Thomas K, Schrupp D, Sajwaj T, Kirby J, Waller E, Schrader S, Falzarano S, … Manis G (2007) Mapping moderate-scale land-cover over very large geographic areas within a collaborative framework: a case study of the Southwest Regional Gap Analysis Project (SWReGAP). Remote Sensing of Environment 08: 59–73. https://doi.org/10.1016/j.rse.2006.11.008

Metzger MJ, Bunce RG, Jongman RH, Sayre R, Trabucco A, Zomer R (2013) A high-resolution bioclimate map of the world: a unifying framework for global biodiversity research and monitoring. Global Ecology and Biogeography 22: 630–638. https://doi.org/10.1111/geb.12022

Muchoney D, Borak J, Chi H, Friedl M, Gopal S, Hodges J, Morrow N, Strahler A (2000) Application of the MODIS global supervised classification model to vegetation and land cover mapping of Central America. International Journal of Remote Sensing 21: 1115–1138. https://doi.org/10.1080/014311600210100

Omernik JM, Griffith GE (2014) Ecoregions of the conterminous United States: evolution of a hierarchical spatial framework. Environmental Management 54: 1249–1266. https://doi.org/10.1007/s00267-014-0364-1

Picotte JJ, Dockter D, Long J, Tolk B, Davidson A, Peterson B (2019) LANDFIRE remap prototype mapping effort: Developing a new framework for mapping vegetation classification, change, and structure. Fire 2: 35. https://doi.org/10.3390/fire2020035

Pliscoff P, Fuentes-Castillo T (2011) Representativeness of terrestrial ecosystems in Chile's protected area system. Environmental Conservation 38: 303–311. https://doi.org/10.1017/S0376892911000208

Radeloff VC, Helmers DP, Kramer HA, Mockrin MH, Alexandre PM, Bar Massada A, Butsic V, Hawbaker TJ, Martinuzzi S, … Stewart SI (2017) The 1990–2010 wildland-urban interface of the conterminous United States – geospatial data, 2nd ed. Forest Service Research Data Archive, Fort Collins, CO, US.

Rollins MG (2009) LANDFIRE: a nationally consistent vegetation, wildland fire, and fuel assessment. International Journal of Wildland Fire 18: 235–249. https://doi.org/10.1071/WF08088

Rollins MG, Frame CK [tech. eds] (2006) The LANDFIRE Prototype Project: nationally consistent and locally relevant geospatial data for wildland fire management Department of Agriculture, Forest Service, Rocky Mountain Research Station [Gen. Tech. Rep. RMRS-GTR-175], Fort Collins, CO, US, 416 pp. https://doi.org/10.2737/RMRS-GTR-175

Rulequest Research Data Mining tools See5 and C5.0 (2012) Data Mining Tools See5 and C5.0. http://www.rulequest.com/see5-info.html [accessed 15 January 2013]

Sahr K (2013) DGGRID, Version 6.1 User documentation for discrete global grid software. http://webpages.sou.edu/~sahrk/dgg/dggrid.v6/dggridManual.v61.pdf [accessed 30 January 2013]

Sano EE, Rosa R, Brito JL, Ferreira LG (2010) Land cover mapping of the tropical savanna region in Brazil. Environmental Monitoring and Assessment 166: 113–124. https://doi.org/10.1007/s10661-009-0988-4

Sayre R, Bow J, Josse C, Sotomayor L, Touval J (2008) Terrestrial ecosystems of South America. In: Campbell J, Jones KB, Smith J, Koeppe M (Eds) North America Land Cover Summit: A Special Issue of the Association of American Geographers, pages from-pages to.

Sayre R, Comer P, Hak J, Josse C, Bow J, Warner H, Larwanou M, Kelbessa E, Bekele T, Waruingi L (2013) A new map of standardized terrestrial ecosystems of Africa. Association of American Geographers, Washington, DC, US, 23 pp.

Sayre R, Karagulle D, Frye C, Boucher T, Wolff NH, Breyer S, Wright D, Martin M, Butler K, Touval J (2020) An assessment of the representation of ecosystems in global protected areas using new maps of World Climate Regions and World Ecosystems. Global Ecology and Conservation 21: e00860. https://doi.org/10.1016/j.gecco.2019.e00860

Soil Survey Staff (2015) Gridded Soil Survey Geographic (gSSURGO) Database for the Conterminous United States. U.S. Department of Agriculture, Natural Resources Conservation Service. https://gdg.sc.egov.usda.gov [accessed Nov 2015]

Stone TA, Schlesinger P, Houghton RA, Woodwell GM (1994) A map of the vegetation of South America based on satellite imagery. Photogrammetric Engineering and Remote Sensing 60: 541–551.

True D, Gordon T, Diamond D (2000) How the size of a sliding window impacts the generation of landforms. Missouri Resources Assessment Partnership. Columbia MO, USA, 22 pp.

U.S. Department of Agriculture (2006) National Agricultural Statistics Service, Cropland Data Layer (CDL). https://nassgeodata.gmu.edu/CropScape [accessed Nov 1, 2016]

U.S. Department of Agriculture (2016) Farm Service Agency, Common Land Unit (CLU) database. Obtained 2006, no longer publicly available.

U.S. Geological Survey (USGS 2016a) National Hydrography Dataset. https://www.usgs.gov/core-science-systems/ngp/national-hydrography/access-national-hydrography-products [accessed Nov 2016]

U.S. Geological Survey (USGS 2016b) 3D Elevation Program (3DEP) 1/3 arc-second resolution Digital Elevation Model. https://www.usgs.gov/core-science-systems/ngp/3dep/data-tools [accessed Nov 2016]

Weiss A (2001) Topographic position and landforms analysis. In: Poster presentation, ESRI user conference, San Diego, CA (Vol. 200).

Weiss GM (1995) Learning with rare cases and small disjuncts. In: Machine learning: Proceedings of the Twelfth International Conference. Morgan Kaufmann, San Francisco, US, 558–565. https://doi.org/10.1016/B978-1-55860-377-6.50075-X

Wickham J, Homer C, Vogelmann J, McKerrow A, Mueller R, Herold N, Coulston J (2014) The multi-resolution land characteristics (MRLC) consortium—20 years of development and integration of USA national land cover data. Remote Sensing 6: 7424–7441. https://doi.org/10.3390/rs6087424

Wickham J, Stehman SV, Sorenson DG, Gass L, Dewitz JA (2021) Thematic accuracy assessment of the NLCD 2016 land cover for the conterminous United States. Remote Sensing of Environment 257: e112357. https://doi.org/10.1016/j.rse.2021.112357

E-mail and ORCID

Patrick J. Comer (Corresponding author, pat_comer@natureserve.org), ORCID: https://orcid.org/0000-0002-5869-2105
Jon C. Hak (jonchak@gmail.com)
Daryn Dockter (ddockter@contractor.usgs.gov)
Jim Smith (jim_smith@tnc.org)

Supplementary material

Supplementary material 1
Spreadsheet includes listing of classification units, hyperlinks to descriptions, and statistics on each type as represented in map products.
Link: https://doi.org/10.3897/VCS.67537.suppl1